

# How do I know which Standard to Use?

Joint Efforts among NCRI, NCBO and NCI CBIIT



U.S. DEPARTMENT  
OF HEALTH AND  
HUMAN SERVICES

National Institutes  
of Health



# Participants

- NCBO
  - Mark Musen
  - Natasha Noy
- NCRI
  - Alan Hogg
  - Stuart Bell
- NCI CBIIT
  - Larry Wright
  - Sherri DeCoronado
- caBIG
  - Stuart Turner
  - Grace Stafford
  - Mike Riben
  - Hua Min
  - Bob Freimuth
  - Brian Davis

# Problem Statement

- How does a naïve (or even experienced) users know which standard to use?
  - Data Type Standard?
  - Modeling Standard?
  - Regulatory Standard?
  - Semantic Standard?
    - Vocabulary Standard?
- We concentrated on Vocabulary Standards: there are 100's of “Vocabulary Standards” to choose from.
- Solution(s) can inform our Orgs

# Goals of Project

- Three Organizations have a shared interest to interoperate: how to represent (vocabulary) Standards to Users
  1. Can we agree on a common way to represent (vocabulary) standards to users?
    - What is the agreed upon metadata on a vocabulary standards that is useful?
  2. Can we make use of community feedback to help users choose standards to use?
    - Amazon type reviews
    - Consumer Reports type of reviews


Leverage each other strengths


# NCBO BioPortal

- An open repository of biomedical ontologies
  - 200 ontologies with 2.5 million terms
  - 4 million mappings between terms
- Community-based metadata and evaluation
  - Ontology **metrics**
    - Computed automatically
  - User **reviews**
    - Submitted by users: star ratings and free-text reviews along different dimensions
  - Other **metadata**
    - “Stamp of approval” from a specific community



# NCBO BioPortal: A Community-Based Ontology Repository


[Browse](#)
[Search](#)
[Projects](#)
[Annotate](#)
[All Mappings](#)
[All Resources Alpha](#)
[Sign In](#)
[Register](#)
[Help/About](#)
[Send](#)

 NCI Thesaurus

## Welcome to the NCBO Bioportal

Use BioPortal to access and share ontologies that are actively used in biomedical communities. You can search for terms in ontologies (try typing "Melanoma" in the "Search all ontologies" box in the left column), browse a list of ontologies in BioPortal (type "NCI Thesaurus" in the "Find an ontology" box in the middle column), search biomedical resources that we automatically annotated with ontology terms (try typing "Melanoma" in the "Search resources" box in the right column). You can [create ontology-based annotations for your own text](#), [link your own project that uses ontologies to the description of those ontologies](#), [find and create relations between terms in different ontologies](#), review and comment on ontologies and their components as you [browse](#) them. [Sign in to BioPortal](#) to submit a new ontology or ontology-based project, provide comments on ontologies or add ontology mappings.

### Search all ontologies


[Advanced Search](#)

### Find an ontology


[Browse Ontologies >](#)

### Search resources


[Advanced Resource Search](#)

### Most Viewed Ontologies (January, 2010)

Ontology	Views
<a href="#">NCI Thesaurus</a>	1 503
<a href="#">Foundational Model of Anatomy</a>	774
<a href="#">RadLex</a>	691
<a href="#">Mouse adult gross anatomy</a>	627
<a href="#">Ontology for Biomedical Investigations</a>	534

### Statistics

Ontologies	195
Terms	1,438,792
Resources Indexed	22

### Latest Notes

[Add Education Service as a Synonym Training Service \(Biomedical Resource Ontology\)](#)  
04/27/10 whetzel  
Based on feedback from Pitt, propose adding Education Service as a Synonym.

---

[Synonym proposal Robotic Arrayer \(Biomedical Resource Ontology\)](#)  
03/23/10 whetzel  
Check with eagle-i for term 'chip spotting device' as a possible synonym.

---

[RE: mis-spelling Physiological Model \(Biomedical Resource Ontology\)](#) 03/14/10 whetzel  
To be fixed in next release.

---

[Term name Molecular Modeling and Classification \(Biomedical Resource Ontology\)](#) 03/14/10 whetzel  
Should this include both modeling and

### Latest Mappings

cellular organisms (NCBI organismal classification) => JERMOntology:organism (SysMO-DB JERM)  
04/14/10 kwoilstencroft

---

[JERMOntology:organism \(SysMO-DB JERM\) => cellular organisms \(NCBI organismal classification\)](#)  
04/14/10 kwoilstencroft

---

[mesenchyme of elbow \(Uber anatomy ontology\) => mesenchyme \(Human developmental anatomy, abstract version\)](#)  
DbXref04/09/10 OBO mapping information

---

hand digit muscle (Uber anatomy ontology) => hand digit muscle (Mouse adult gross anatomy)  
DbXref04/09/10 OBO mapping information

---

[mesenchyme of carpal region \(Uber anatomy ontology\) => mesenchyme \(Human developmental anatomy, abstract version\)](#)  
DbXref04/09/10 OBO mapping information

<http://bioportal.bioontology.org>

# NCRI Cancer InfoMatrix

- Consistent promotion of standards
- An easy place for non-ontology specialists to find information on what standards to use in their domain
- Highlighting that while vocabularies/ontologies are critical, other factors have equal importance
  - Data interchange formats
  - Reporting guidelines
  - Tools that support standards

# NCRI: Cancer InfoMatrix

The screenshot shows the NCRI Oncology Information Exchange website in a Windows Internet Explorer browser. The page title is "NCRI ONCOLOGY INFORMATION EXCHANGE". The main navigation menu includes HOME, NEWS & EVENTS, Cancer InfoMatrix (highlighted), RESOURCE CATALOGUE, QUICK SEARCH, DOCUMENT REPOSITORY, and LINKS. A login section is present with fields for Email Address and Password, and a Login button. Below the login section are links for "Become a registered user", "Forgotten my password", and "Problem accessing my account". The Cancer InfoMatrix section is titled "Cancer InfoMatrix" and provides information about the CIM, which is a grid of data standards development activities. The grid has columns for DNA, RNA, Proteins, Metabolites, Tissue, Models, Systems, Images, Clinical, and Population. The rows are Data Reporting Guidelines, Ontologies & Controlled Vocabularies, Data Exchange Formats, and Standards Tools. Each cell in the grid contains a battery icon representing the degree of progress. A legend at the bottom indicates the activity level: High (green), Medium (yellow), and Low (red). The footer contains links for Terms of Use, Contact Us, Privacy, Site Map, and copyright information for NCRI 2010.

NCRI ONCOLOGY INFORMATION EXCHANGE

Find out more | FAQ | Help

HOME  
NEWS & EVENTS  
**Cancer InfoMatrix**  
RESOURCE CATALOGUE  
QUICK SEARCH  
DOCUMENT REPOSITORY  
LINKS

Email Address:  
Password:  
Login

[Become a registered user](#)  
[Forgotten my password](#)  
[Problem accessing my account](#)

NCRI INFORMATICS INITIATIVE  
NHS  
National Institute for Health Research  
caBIG

Cancer InfoMatrix

The Cancer InfoMatrix (CIM) provides information about work being conducted to support good data management in cancer research through the development of [data standards](#). Biomedical domains are arranged across the horizontal axis of the CIM while on the vertical axis are broad areas of data management that enable data sharing, access and exchange. Clicking on a cell within the CIM will take you to a list describing each initiative in that area as well as other relevant information.

The cells in the Cancer InfoMatrix are colour coded using a "traffic-light-battery system" to indicate the degree of progress in each area:

You can make suggestions for additions to the CancerInfo Matrix by [contacting us](#).

	DNA	RNA	Proteins	Metabolites	Tissue	Models	Systems	Images	Clinical	Population
Data Reporting Guidelines	High	High	High	High	High	Low	High	Low	High	High
Ontologies & Controlled Vocabularies	High	High	High	High	High	High	High	High	High	High
Data Exchange Formats	High	High	High	High	High	High	High	High	High	Low
Standards Tools	High	High	High	High	High	High	High	High	High	High

Data Standards Development Activity: High Medium Low

Terms of Use | Contact Us | Privacy | Site Map | © NCRI 2010 | v1.4.3

Designed & maintained by NCRI INFORMATICS INITIATIVE

524

<http://www.ncri-onix.org.uk/portal/#S24>



# NCI Enterprise Vocabulary Service & caBIG Vocabulary Knowledge Center

- Disseminate information about tools and standard terminologies to user communities
- Support for NCI EVS Semantic Infrastructure
  - Terminology services, authoring, and editing
- Support for NCI EVS Terminologies
  - NCI Thesaurus (75,000 concepts)
  - NCI Metathesaurus (1.4 M concepts mapped to 3.6 M terms with 20 M relationships)
  - NCI Term Browser (for all hosted terminologies)
- Encourage community engagement/feedback

# NCI CBIT and caBIG

The screenshot displays the NCI caBIG Knowledge Center website. The header includes the National Cancer Institute logo and name, the U.S. National Institutes of Health logo, and the website URL www.cancer.gov. Below the header is a navigation bar with links to Home, Knowledge Centers, Discussion Forums, Bugs/Feature Requests, and Development Code Repository. A search bar is located on the left side of the page. The main content area is titled "Terminologies" and includes a sub-header "Main Page > Terminologies". A sidebar on the left lists various content categories and tools. The main content area features a list of terminologies, including Gene Ontology (GO), Logical Observation Identifier Names and Codes (LOINC), NCI Thesaurus (NCIt), RadLex (RID), Systematized Nomenclature of Medicine-Clinical Terms (SNOMED CT), and Medical Dictionary for Regulatory Activities. A callout box highlights the NCI Thesaurus, providing a definition, a link to the discussion page, and a list of links for further information.

**NCI Thesaurus**

[Main Page > Terminologies > NCI Thesaurus](#)

**What is the NCI Thesaurus?**

The NCI Thesaurus is a reference terminology and biomedical ontology used in a growing number of NCI and other systems. It covers vocabulary for clinical care, translational and basic research, and public information and administrative activities. The NCI Thesaurus provides definitions, synonyms, and other information on nearly 10,000 cancers and related diseases, 8,000 single agents and combination therapies, and a wide range of other topics related to cancer and biomedical research. It is maintained by a multidisciplinary team of editors, who add about 900 new entries each month and is published monthly by NCI.

For assistance with the use of this terminology please send a note to [NCIThesaurus Help](#)

**Discuss NCI Thesaurus**

Please go to the [NCI Thesaurus discussion page](#) to leave comments regarding the NCI Thesaurus.

Examples of comments may include:

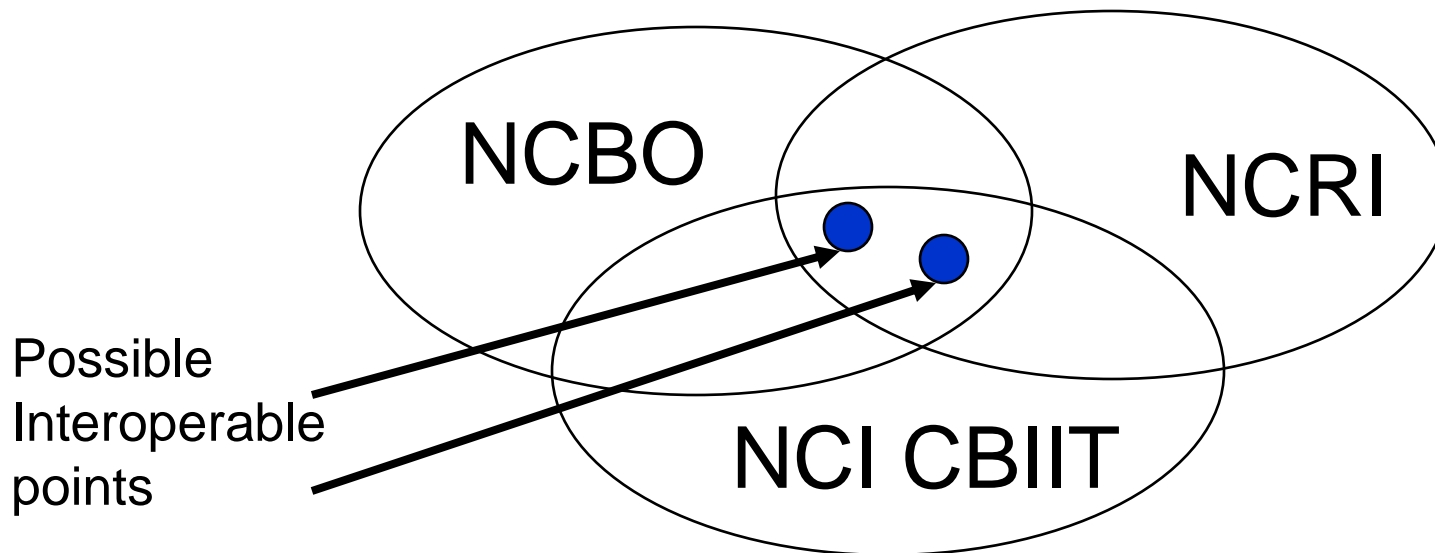
- How you are using this terminology
- Why you decided to use (or not use) this terminology
- Strengths or limitations of the terminology
- Comparisons to other similar terminologies

**Links**

- [NCI Thesaurus FAQ](#)
- [NCI Thesaurus Wiki Home Page](#)
- [NCI Thesaurus Download](#)
- [NCI Thesaurus Browser](#)

# Overlap of Goals among three Organizations

- Advertise useful metadata about vocabulary standards to users
- Community rating system on vocabulary standards



# What metadata about Vocabularies is most useful to users? OMV(plus)

- Multiple possibilities for metadata “standard” for describing vocabularies for users
- The Ontology Metadata Vocabulary (OMV) seems likely candidate
- NCBO has made extensions
- Testing now in NCRI and NCI CBIIT to determine if appropriate for users

# OMV: Ontology Metadata Vocabulary

- OMV Consortia:
  - University of Bremen (TZI)
  - Universidad Politecnica de Madrid (UPM)
  - University of Karlsruhe (UKARL)
  - Stanford Center for Biomedical Informatics Research (BMIR)
- Question:
  - If we use and adapt, how do we interact with OMV Consortia?



# Some OMV properties describing an ontology (properties on *OMV:Ontology*)

- *OMV:acronym*
- *OMV:name*
- *OMV:URI*
- *OMV:naturalLanguage*
- *OMV:creationDate*
- *OMV:modificationDate*
- *OMV:description*
- *OMV:designedForOntologyTask*
- *OMV:documentation*
- *OMV:endorsedBy*
- *OMV:hasContributor*
- *OMV:hasCreator*
- *OMV:hasDomain*
- *OMV:status*
- *OMV:cointainsABox*,  
*OMV:containsTBox*
- *OMV:expressiveness*
- *OMV:hasFormalityLevel*
- *OMV:hasLicense*
- *OMV:keywords*
- *OMV:keyClasses*
- ***OMV:knownUsage***
- *OMV:isOfType*
- *OMV:usedOntologyEngineeringTool*
- *OMV:usedKnowledgeRepresentation  
Paradigm*
- *OMV:numberOfClasses*
- *OMV:numberOfIndividuals*
- *OMV:numberOfAxioms*
- *OMV:numberOfProperties*

# NCBO Extensions

- *administeredBy*
- *hasContactEmail*
- *hasContactName*
- *uploadDate*
- *id*
- *internalVersionNumber*
- *preferredNameProperty*
- *synonymProperty*
- *documentationProperty*
- *authorProperty*
- *codingScheme*
- *fileNames*
- *filePath*
- *hasView*
- *isVersionOfVirtualOntology*

# Next steps/value to Organizations

- Test OMV(plus): NCRI and NCI CBIIT on existing Vocabulary Standards
- Analyze “community review” representations
- Process that supports harmonization (among the three organization).
- Prototype for Sharing (operationalizing)

